树立数据思维，建设新闻出版大数据体系

摘　要：实际上，早在20世纪80年代，美国未来学家Alvin Toffler就在他的《第三次浪潮》中对“大数据”这一概念赞不绝口，他将其称为“第三次浪潮中的华彩乐章”，早早就面向世界展示了大数据的内在潜力。进入21世纪，大数据技术被全球所广泛应用，成为了信息技术行业中的至宝，树立数据思维，建设行业发展大数据体系也成为行业未来的重中之重。本文主要以新闻出版行业为例，探讨了在新闻出版行业中树立大数据思维，建设行业大数据体系的具体思路、建设方案，并给出相关发展建议。

关键词：大数据体系；数据思维；新闻出版行业；建设思路；方案；建议

中图分类号：G641　　**文献标识码：**A

文章编号：1671-0134（2019）12-066-03　　**DOI：**10.19483/j.cnki.11-4653/n.2019.12.018

本文著录格式：吴翠姣，张宇．树立数据思维，建设新闻出版大数据体系[J]. 中国传媒科技，2019（12）：66-68.

文 / 吴翠姣　张宇

2012年，美国白宫科技政策办公室首次发布了有关大数据技术应用的相关计划《大数据研究和发展计划》，并专门为此成立了大数据高级指导小组，将大数据战略真正上升为国家级发展战略。2014年，我国的《政府工作报告》中也首次提出了“大数据”这一概念，并在2015年的《关于运用大数据加强对市场主体服务和监管的若干意见》中明确了国家实施政府数据资源、构建公共服务大数据工程、优化大数据。我国希望在未来完成的3件事：政府数据资源政策的实施、公共服务大数据工程的建设、大数据优化关键技术及相关产品的相关工程建设内容，构建了真正的大数据产业支撑能力体系。4年来，我国纷纷建立了多项大数据工程项目及相关安全保障工程项目，目前，我国已经推出了“互联网+大数据”行动计划，希望实施网络强国战略，开启新经济时代下大数据建设的新篇章。

1. 大数据体系中的创新数据思维

在信息如洪流的21世纪，大数据技术体系的出现颠覆了时代发展进程，改变了社会发展节奏，围绕大数据技术所构建的大数据体系希望革新人类未来发展思维，这一点在注重社会交流的新闻出版行业尤其突出。目前的新闻出版行业从业人员必须更新自己的世界价值观，深入自身所从事的社会活动与工作岗位，获取利用各方面大数据信息内容，结合生产加工与传播消费，凸显大数据技术应用重要价值，在构建大数据体系的背景下不断创新数据思维，引导社会生产生活。

1.1 新闻出版行业需要拥有的3种创新数据思维

有一句谚语：“除了上帝，任何人都要相信并应用数据，用数据说话。”这一理念在我国的《促进大数据发展行动纲要》中就提出了这一观点，即“建立用大数据说话，用数据决策、管理、创新中同样也提到了相同观点，要建立围绕大数据技术的、用大数据说话、决策、创新、管理的一套完整的、功能丰富的全新大数据体系。”这一管理体系的核心自然就是大数据，包括大数据中所传承的数据思维与数据文化内容。在新闻出版行业中，大数据就是要上升到文化这一高度，这是因为它必须追求一种理性思维本质和科学精神，做到新闻说话实事求是，正确引导社会发展舆论，用数据证明事实，这是新闻出版行业需要遵循的最基本原则。奥地利大数据专家Viktor Mayer-Schönberger就在他的《大数据时代》中阐述了具有洞察性质与前卫理念的3种大数据创新思维，它们分别为整体思维、多样性思维及相关思维。但实际上，对于新闻出版行业而言，可以将这3种思维理解为价值思维、应用思维及共享思维。

1.2 价值思维

大数据对于新闻出版行业发展而言就是一种无形的资产，所以，行业一直遵循数据即财富的基本观念，大数据的核心就是这一价值思维。在新闻出版社中，他们所拥有的价值思维包括了数据管理、数据交易、数据交易行为以及用户数据，特别是涵盖了数据管理、交易、交易行为，以及用户数据等相关内容，尤其是传统新闻出版社需要意识到一点，用户对于他们而言就是一种不可被忽略的隐形财富。早在2014年，Facebook宣布用160亿美元收购Whats App时，他们就证明了这一点，用有形资产换取大量的无形资产（用户）是非常值得的，这也是Facebook的成功之道。

1.3 应用思维

在21世纪，拥有大数据的人就拥有了未来。此话有待商榷，因为大数据能够帮助行业企业占据制高点，取得主动权，这就说明了数据的价值在于应用，而不在于拥有。新闻出版行业需要大量与社会进行交流沟通，他们需要利用数据驱动自身内部的所有业务流程与产品生产研发过程，所以说，新闻出版行业需要利用数据驱动业务，用业务产生价值，这也是传统新闻出版社希望发展转型的原因之一。相比于生产流程，他们更需要数据，

因为生产流程只能给予他们固定的定量思维，但是数据却能为他们提供流动的定性发展思维，例如，在数据决策、印数决策、选题策划等方面都需要数据作为支撑，通过架构一套完整的大数据平台体系、深入构建全维度记录体系、优化存储于描述数据关系架构商务大数据平台，建立全维度记录、存储与描述数据体系，真正深入到数据关联、数据挖掘技术应用过程中，对数据实施相关预测与推荐过程，这是为了有效解决存在于传统新闻出版行业中的缺陷问题，为新闻出版产品与服务内容提供数据化服务条件进行预测与推荐，最后解决传统新闻出版行业中所存在的种种痛点，最后实现新闻出版产品与服务的数据化。

1.4 共享思维

目前的许多新闻出版社中存在着严重的数据孤岛现象，造成这一问题的原因是出版社、编辑室、作者三者之间无法实现信息共享，因此，出版社本身可能会陷入到孤岛效应危机中。再者就是传统新闻出版社是缺乏大数据平台做支撑的，这就导致他们在业务发展过程中无法实现行业协同，无法实现对数据的共享过程。按照上文所谈到的“共享即为使用而并非拥有”这一概念来看，行业内部数据流动性并不强，无法与外部数据相互融合，最终成为死数据。新闻出版社需要解决的问题有二：第一就是合理平衡数据的共享关系，优化共享效率；第二则是建立基于利益平衡机制的共享平台，这是传统新闻出版社走向未来转型发展路径的关键。[1]

2. 新闻出版行业需要解决的现实发展问题

传统新闻出版行业是存在诸多发展问题的，它们在某种程度上严重影响了产业健康向前发展进程。首先是传统行业中存在严重的产能过剩、供需不匹配等问题，客观且严重地影响了新闻出版行业发展建设所存在的各种问题，导致传统行业中出现了包括产能过剩、供需无法匹配等现实问题，它们为行业带来的风险极大，所以目前行业是希望追求绿色发展前景的。从 2014 年大数据概念出现后看，传统新闻出版社还未改变自身发展格局，当时全国的新闻出版刊物纯销售额只有 780 亿元，已经不可避免地形成倒挂发展之势。与此同时，发行市场中所存在的群雄并起、诸侯割据问题也导致全国性发行中的严重缺乏，统一开放、竞争有序的全国大市场成为一种奢望，这直接导致我国出版传媒集团难以做大做强。再一点就是出版发行生态链目前处于严重的畸形发展态势，它所运行的寄销制导致其无条件退货、账期较长、回款困难、诚信丧失等问题的出现，图书单品种核算问题。客观讲，目前的泛行业发展是追求粗放式管理方式的，如果无法从本质上思考并解决相关桎梏问题，就可能会导致大数据技术体系与行业大数据平台、信用信息平台发展建设陷入被动。而目前，争取从产业链的上、中、下游实施全方位精确化管理，确保数据驱动选题策划到位，同时提供印刷决策内容，深度、精准分析用户行为与营销行为，同时构建个性化发展机制，为出版社重塑生产、销售与管理模式创造无限可能性是很有必要的，新闻出版行业企业在发展转型进程中需要准确把握这一点。[2]

3. 新闻出版行业中的大数据内容与大数据体系构建方法

3.1 新闻出版行业中的大数据内容

大数据是目前新闻出版行业快速发展的关键，它迎合了信息化时代发展建设标准，但就目前来看，新闻出版行业中对大数据及其相关技术的应用也是仁者见仁、智者见智的。就目前出版社中的大数据内容来看，它的分类方法众多，其中主要结合业务层面进行划分，新闻出版社的大数据内容主要包括 6 类：第一是机构数据，例如新闻出版政府机构、行业协会、事业单位、出版社、报社、发行集团等，它还包含了企业的出版收入与营销利润；第二是人员数据内容，其中就包括了公务员、编辑记者、专家、新闻出版社以及行业中所有从业人员的基础数据等；第三是产品数据，它主要包括图书报刊数据、音像数据、网络出版物数据等，同时也包含了作者的元数据、印刷数据、版权数据等；第四是政务数据，其中包含了政府的综合办公数据、市场监管数据、公共服务过程产生的数据以及人事信息、财务信息数据等；第五是商务数据，其中包括了新闻出版社的信息企业主体市场交易数据、合同数据以及版权数据等；第六是内容数据，其中包含了信息知识数据，这部分数据包含 4 种类型，其中包括了事实知识、技能知识、原理知识与人力知识，它形成了丰富的大数据库平台，被视为新闻出版行业中的特有数据宝藏内容，也成为了知识服务的重要基础。[3]

3.2 新闻出版行业中的大数据体系构建过程

3.2.1 数据采集

数据采集是目前新闻出版行业生产发展转型中的最重要因素，它希望结合数据类型内容中的内容数据、市场数据与用户数据构建新的数据发展业态，强调大数据技术内容应用的交互性内涵。在整个过程中，它专门围绕出版行业核心业务内容建设有价值的、有时效性的信息发展体系，将数据采集与出版机构内容创造所获得的回馈数据融入到实际生产进程中。目前，我国确实存在大量的权威专业新闻出版社，它们拥有丰富的专业内容资源，且在出版业务转型升级过程中更加追求数据内容的资源化与碎片化发展，强调利用大数据提供丰富资源内容，确保有效实现行业数据衔接调整，解决存在于企业中的某些“数据孤岛”效应，寻求行业产业快速发展，强化行业数据发展规范标准，满足产业发展需求，建立有针对性的大数据服务体系。另一方面，需要满足新闻出版行业的统一规范化数据采集标准要求，不遗余力地优化大数据采集平台，强化行数据发展规范标准。就目前来看，基于大数据分析与应用的全数据资源体系构建是有必要的，它能够在一定程度上提升大数据应用的时效性与准确性。[4]

3.2.2 数据标引

新闻出版社在大数据技术支持下可实现对海量数据的采集，再结合所获得的数据实施标引工作。在整个过程中需要对大数据体系中的预测功能、预警功能进行分

析应用，提升数据创新关键作用。就我国整个出版行业发展现状来看，行业企业还应当更多结合、利用数据标引技术内容实现对知识与行业应用内容的有效标引，建立一套知识标引学科研究体系，这也能为后续发展新闻出版社的知识服务内容奠定良性基础。[5]

3.2.3 数据计算

新闻出版社的数据计算工作内容包括了云计算与统计学分析，二者可实现兼容并包，同步实施。其中，首先分析云计算内容，它所蕴藏的大数据内容是海量的，且这些海量数据能够与标引数据内容相互结合，实现对数据云计算处理过程优化，建立基于处理结果的二次数据研究体系。在这里，云计技术与数据计算过程中都会提供丰富且多元化的动态计算内容，实现对技术内容的扩展与虚化，确保虚拟化信息资源内容实现全面调整，真正被纳入到网络链接中，真正满足资源统一化管理与智能化调度过程。客观讲，它所构建的就是一套全新的新闻出版社数据资源池，其池中不断面向某些有需求的用户提供不同的资源服务内容。

而在统计学分析方面，则主要结合计算机技术快速发展进程对统计数据进行相应搜集与处理，同时满足数据分析与存储发展需求。考虑到一般计算机系统是无法对数据所反映的客观复杂规律进行有效厘定的，因此，需要在判断未来发展趋势过程提出决策方案内容，对数据内容进行二次标引与计算，满足数据统计学分析条件。在该过程中，还需要对数据的不确定性进行量化分析，基于数据信息规律提取数据内容，找到数据标引与统计计算的最优化方案。[6]

3.2.4 数据建模

在数据建模过程中，需要结合不同应用范围对学科体系建模与行业应用建模进行分析。在该过程中，学科体系建模主要依赖于当前已有学科体系，它拥有相当成熟的理论知识体系，可根据相关关系构建内在模型，选择合理化数据建模方案。目前的新闻出版行业中的大数据建模结构相当成熟，可作为行业发展应用建模使用，它的建模范围涉猎各个行业发展领域，可结合行业职业基础现状发展扩大数据建模范围，这也意味着新闻出版企业业务范围的扩大。而且数据建模是个性化的，它能够针对用户的特定化需求构建相应大数据模型，主要针对数据建模的发展领域。[7]

3.2.5 数据服务

在数据服务方面，它主要结合数据采集、标引、统计、建模等环节展开数据服务体系设计，争取为所有目标客户提供丰富多彩的大数据服务内容，充分考量新闻出版行业中的审计发展内容有效转型与优化。具体来讲，就是合理利用大数据技术全方位推动出版行业机构内容从出版到知识服务传播方面有效转型，例如：从人工智能向发展机制的转型，从传统业务门类面向大数据驱动方向的转型，全方位提升新闻出版社机构的整体业务运营与发展管理效率。

具体来讲，目前的新闻出版社应该追求利用大数据技术优化个性化知识服务内容，从不同渠道整合数据资源，确保数字化整理背景下对知识服务平台的有效优化，建立围绕用户展开服务的检索与阅读平台、热点分析平台、数据挖掘分析平台以及知识关联服务平台。与此同时，要构建知识图谱，为用户提供数字化、知识化与智能化服务项目。而目前的新闻出版社是追求数字化出版业务发展的，因此，基于大数据的个性化用户检索、浏览终端行为实施都是可以理解的，它们都能精确掌握当前的用户实际需求规律，并为用户未来的知识需求定制个性化方案。[8]

总结

未来大数据模型标准化建设与应用应当成为当前新闻出版行业发展的基础关键。如我国的国标委就已经成立了“全国信标委大数据标准工作组”，负责制定一套完善的国家大数据领域标准体系，为新闻出版结合大数据技术过程采集多方面资源数据内容，实现信息数据有效对接。所以，在未来，新闻出版行业还应该建立基于大数据共享交换机制的信息化业务平台，加速建设大数据重大工程，为新闻出版社大数据体系的建设与业务创新优化奠定基础。媒

参考文献

[1] 刘成勇 . 树立数据思维，建设新闻出版大数据体系 [J]. 出版参考，2016（7）：4-8.

[2] 曾文，徐红姣，车尧，等 . 基于图书出版行业大数据的选题决策分析模型研究 [J]. 情报学报，2018，37（8）：813-821.

[3] 田军 . 新闻出版业大数据运用的新业态研究 [J]. 传播与版权，2017（8）：47-49.

[4] 王欣欣 . 国内新闻出版产业大数据应用困境分析 [J]. 网络安全技术与应用，2016(10):120-121.

[5] 石佳靓，吴鼎 . 国内新闻出版产业大数据应用障碍分析 [J]. 中国出版，2015（23）：29-31.

[6] 张新新 . 新闻出版业大数据应用的思索与展望 [J]. 科技与出版，2016（1）：3-8.

[7] 无 . 新闻出版大数据用户行为跟踪与分析实验室成立 [J]. 新闻世界，2017（4）.

[8] 王婷婷 . 大数据时代传统出版业的转型思考 [J]. 商业经济，2017（4）.

（**作者单位：**中国协和医科大学出版社）